Available online at www.sciencedirect.com
ScienceDirect
Scientia Horticulturae 115 (2007) 19–26
SCIENTIA HORTICULTURAE
www.elsevier.com/locate/scihorti

Usefulness of WRKY gene-derived markers for assessing genetic population structure: An example with Florida coconut cultivars

Margarita Mauro-Herrera, Alan W. Meerow *, James W. Borrone, David N. Kuhn, Raymond J. Schnell

USDA-ARS-SHRS, National Germplasm Repository, 13601 Old Cutler Road, Miami, FL 33158, USA

Received 23 February 2007; received in revised form 6 July 2007; accepted 10 July 2007

Abstract

We previously analyzed genetic diversity and population structure of eight cultivars within Florida coconut (*Cocos nucifera* L.) germplasm using 15 microsatellite (simple sequence repeat, SSR) markers. Here we report on the analysis of the same genotypes using 13 markers derived from WRKY sequences containing single nucleotide polymorphisms (SNP) and one microsatellite. WRKY transcription factors are unique to plants and are involved in numerous vital processes including disease resistance. Our objective was to determine the value of this type of gene to assess the genetic diversity of this species. Overall, the WRKY results are similar to those with microsatellite markers. Despite the lower number of alleles identified with the WRKY-derived markers (37 versus 67 total alleles identified with the SSR markers), individuals of cultivars 'Red Malayan Dwarf', 'Fiji Dwarf' ('Niu Leka') and 'Red Spicata' were clearly clustered, as in the previous analysis. However, individuals of 'Green Malayan Dwarf' and 'Yellow Malayan Dwarf' cultivars were resolved with other varieties, perhaps due to selective forces operating on these functional genes. Most samples of the 'Atlantic Tall' and 'Panama Tall' cultivars clustered together as in the previous microsatellite study. We discuss the utility of WRKY-derived markers for assessing the genetic diversity of coconut, and their possible value in the study of other species. Published by Elsevier B.V.

Keywords: Coconut; *Cocos nucifera*; Genetic diversity; SNP; WRKY genes

1. Introduction

Cocos nucifera L., the coconut palm, possesses significant levels of genetic diversity (Lebrun et al., 2005), often not perceptible at the morphological level but easily explained by the diverse habitats where it is found (Foale, 2005). It grows throughout the world in tropical and subtropical environments, and is commonly associated with tropical coastal habitats. In many developing countries it has several important agronomic and subsistence uses (Harries et al., 2004), while in the United States it is used primarily as an ornamental. Coconut cultivars are generally classified into tall and dwarf types. The tall types are primarily outcrossing while the dwarf types are predominantly self-pollinated (Liyanage, 1949). This difference in breeding habit has lead to higher levels of genetic diversity in the tall types, which contain around 60% of the total diversity (Rao et al., 2005). Several marker types have been used in the genetic analysis of coconut varieties (Rohde et al., 1995; Ashburner et al., 1997; Lebrun et al., 1998; Perera et al., 1998, 1999). We were interested in the development of markers that could be useful for breeding and selection purposes. Therefore, we focused on WRKY genes to develop our markers, and assessed their ability to establish genetic relationships and levels of diversity in a group of cultivars previously evaluated with microsatellite markers.

The WRKY transcription factor family was first described by Eulgem et al. (2000). These transcription factors have a well-conserved amino acid sequence, the WRKYGQK domain, from which their name originated. WRKY transcription factors are involved in numerous biological functions in plants. Several WRKY genes have been identified as important components in the genetic control of defense response mechanisms in a number of plant species (Eulgem et al., 1999; Yu et al., 2001; Ulker and Somssich, 2004; Ryu et al., 2006). For instance, in *Arabidopsis*, 49 of 72 WRKY genes examined were differentially regulated in response to bacterial infection or salicylic acid treatment (Dong et al., 2003). WRKY genes are also induced in the response to abiotic stresses like cold and

* Corresponding author. Tel.: +1 305 254 3635; fax: +1 305 969 6410.
E-mail address: alan.meerow@ars.usda.gov (A.W. Meerow).

0304-4238/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.scienta.2007.07.005

drought (Rizhsky et al., 2002). In addition, this type of transcription factor also participates in senescence (Robatzek and Somssich, 2001), trichome and seed coat development (Johnson et al., 2002) and regulation of seed size (Luo et al., 2005). WRKY genes are extensively represented in plant genomes, with numbers as high as 70 for *Arabidopsis* (Eulgem et al., 2000; Dong et al., 2003) and up to 109 for rice (Zhang and Wang, 2005). WRKY transcription factors seem to be randomly distributed across the *Arabidopsis* genome (Eulgem et al., 2000). However, in rice their distribution is less even with over 20 WRKY genes on chromosome 1 and 2 on chromosome 10. Some of these loci appear to be unique genes while others are likely paralogues (Xie et al., 2005; Zhang and Wang, 2005).

The WRKY conserved domains were useful in the development of molecular markers in *Theobroma cacao* L. (Borrone et al., 2004). Using a similar approach we developed 13 WRKY-derived markers in coconut (Mauro-Herrera et al., 2006). We evaluated these 13 WRKY-derived markers with 110 coconut genotypes previously characterized with 15 microsatellite (simple sequence repeat, SSR) markers by Meerow et al. (2003), and present the results in this paper.

2. Materials and methods

2.1. Plant material

We used the same plant material as in Meerow et al. (2003). The 110 genotypes evaluated were: nine 'Atlantic Tall' genotypes, 36 'Fiji Dwarfs' (also known as 'Niu Leka'), nine 'Green Malayan Dwarfs' (plus two off-types), ten 'Green Niños', six 'Panama Talls', 19 'Red Malayan Dwarfs' (plus one off-type), 11 'Red Spicatas' and five 'Yellow Malayan Dwarfs'. One 'Maypan', a hybrid between a 'Panama Tall' and a 'Malayan Dwarf' variety, and an undetermined tall were also included. The plant material was sampled from two locations in the state of Florida, USA. The full description and provenance of the plant material was reported in Meerow et al. (2003).

2.2. Data collection

The development of ten of the WRKY-derived markers used in this study was reported in Mauro-Herrera et al. (2006). From three additional WRKY sequences we generated three more markers (Table 1). All 13 WRKY-derived primer pairs were used in this study. Polymerase chain reactions were formulated as previously reported (Mauro-Herrera et al., 2006). For 12 of the WRKY-derived markers we used single-strand conformation polymorphism (SSCP) for allele detection at 19 °C and 25 °C, as in Kuhn et al. (2005). Microsatellite analysis for CnWRKY-01 was accomplished as described in Mauro-Herrera et al. (2006). All 13 WRKY-derived markers were co-dominant.

2.3. Data analysis

The program Convert (Glaubitz, 2004) was used to generate the formatted files for several of the programs used in the data analysis, and to produce the table of unique alleles and their frequencies (Table 3). The population genetic structure of all 110 genotypes was estimated based on the WRKY data (this paper) and the microsatellite data (from Meerow et al., 2003) using the Bayesian model-based clustering program STRUCTURE (Pritchard et al., 2000; Falush et al., 2003). All of the other analyses presented here were exclusively performed on the WRKY data, as the same analyses on SSRs were already presented in Meerow et al. (2003). STRUCTURE infers population structure based on distinct allele frequencies, and determines individual membership and possible admixed/hybrid individuals (Pritchard et al., 2000; Falush et al., 2003). The number of populations/clusters (K) was determined by running seven replicate simulations for each K value ranging from 3 to 16. For all runs we used the admixture model with correlated allele frequencies, 7×10^4 iterations for the burn-in period and 10^5 iterations for the running period. The final analysis, with the determined K based on the estimates for the natural logarithm of the probability of the data (Pritchard et al., 2000; Falush et al., 2003), was generated using 10^5 iterations for the burn-in period and 10^6 iterations for the running period. The program Distruct (Rosenberg, 2004) was used to produce the graphic displays with the STRUCTURE results.

For further analyses, and after comparing the STRUCTURE results of both marker systems, we removed the same off-types and putative hybrids as in Meerow et al. (2003), so a fair comparison of the two studies could be made. We used GDA 1.1 (Lewis and Zaykin, 2001) to calculate some of the descriptive statistics (allele number, sample size, proportion of polymorphic loci, expected and observed heterozygosity, fixation index; Tables 2–4). Gene diversity was calculated with FSTAT (Goudet, 2002) using Nei's unbiased estimator (Nei, 1987): $H_{sk} = (n_k/n_k - 1)\left(1 - \sum p_{ik}^2 - (H_{ok}/2n_k)\right)$, where n_k is the size of sample k, p_{ik} is the frequency of allele A_i in sample k, and H_{ok} is the observed proportion of heterozygotes in sample k. We used the same genetic distance estimates as in Meerow et al. (2003), i.e. modified Rogers' distance (Wright's, 1978) and Nei's (1978) unbiased distance, calculated with the program Tools for Population Genetic Analyses (TFPGA; Miller, 1997). Genetic distances between

Table 1
Information for the three additional WRKY-derived markers introduced in this paper

Locus	Primer sequence (5′-3′)	Ta[a] (°C)	Size (bp)	Polymorphism type
CnWRKY-04	F: GCACCCAGTTTCCAAATCTTTCCA R: CACCCTCTTCTTCACCGAGCA	52	215	SNP: C/T
CnWRKY-09	F: TGGAGGAAGTATGGGCAGAAGG R: GAAGCAAGCAATCAACCAAAGAGC	52	195	SNP: G/A
CnWRKY-14	F: CACTGTCATTTAGTCCCGAGCC R: CTCCGCACACCTTTGGACCT	54	142	Possibly SNPs, detected via SSCP

[a] Annealing temperature for PCR.

Table 2
Information for the 13 WRKY-derived markers used in the genetic analysis of the *Cocos nucifera* cultivars

Locus	No. of alleles	He[a]	Ho[b]	f[c]	Gene diversity
CnWRKY-01	3	0.508	0.191	0.625	0.520
CnWRKY-02	2	0.058	0.036	0.387	0.075
CnWRKY-03	3	0.496	0.294	0.408	0.502
CnWRKY-04	3	0.231	0.197	0.147	0.272
CnWRKY-05	3	0.540	0.157	0.710	0.542
CnWRKY-06	2	0.118	0.057	0.519	0.124
CnWRKY-09	2	0.503	0.190	0.624	0.501
CnWRKY-10	4	0.361	0.156	0.571	0.364
CnWRKY-13	4	0.406	0.106	0.740	0.404
CnWRKY-14	3	0.447	0.212	0.528	0.466
CnWRKY-16	3	0.149	0.101	0.322	0.173
CnWRKY-19	2	0.136	0.011	0.918	0.138
CnWRKY-21	3	0.422	0.188	0.556	0.400

[a] Expected heterozygosity.
[b] Observed heterozygosity.
[c] Fixation index.

populations computed with each marker system were compared via Mantel's tests (Mantel, 1967) with 10,000 permutations. Cluster analysis was performed using POPULATIONS (Langella, 2002), with the neighbor-joining method (NJ, Saitou and Nei, 1987) and the unweighted pair-group method using arithmetic averages (UPGMA, Sneath and Sokal, 1973). The Multi-Variate Statistical Package (MVSP, Kovach Computing Services, Anglesey, Wales) was used for principal coordinate analysis using the Rogers' genetic distance estimates between individuals.

3. Results

The STRUCTURE simulation analyses with the WRKY data identified six populations ($K = 6$) in all 110 genotypes evaluated; five of them could unambiguously be associated with a particular cultivar (Fig. 1A). Most of the genotypes of 'Atlantic Tall', 'Fiji Dwarf', 'Red Malayan Dwarf', 'Green Niño' and 'Red Spicata' resolved consistent membership in their corresponding cultivar groups. The hybrid derivation of 'Maypan' from a 'Malayan Dwarf' and a tall coconut is confirmed by the STRUCTURE results. Likewise the genotypic partitioning of the off-types (e.g. otGoMD1 and otGrMD1)

Table 3
Unique alleles identified by the WRKY-derived markers, and their frequency in the eight *C. nucifera* cultivars studied

Cultivar	Locus	Allele	Frequency
'Fiji Dwarf'	CnWRKY-01	207	0.307
'Fiji Dwarf'	CnWRKY-05	292	0.129
'Fiji Dwarf'	CnWRKY-10	427	0.333
'Fiji Dwarf'	CnWRKY-10	450	0.074
'Fiji Dwarf'	CnWRKY-13	196	0.769
'Fiji Dwarf'	CnWRKY-14	282	0.305
'Panama Tall'	CnWRKY-16	496	0.500
'Red Spicata'	CnWRKY-19	324	0.812
'Yellow Malayan Dwarf'	CnWRKY-13	194	0.100

Table 4
Descriptive statistics for 110 coconut (*Cocos nucifera* L.) genotypes as determined with 13 WRKY-derived markers and 13 SSR markers (from Meerow et al. (2003))

Data source	Sample size[a]	P[b]	A[c]	Ap[d]	He[e]	Ho[f]	F[g]
WRKY data of all 110 individuals	100.92	1	2.84	2.84	0.344	0.184	0.465
SSR data of all 110 individuals	104.54	1	4.31	4.31	0.555	0.300	0.460

[a] Average sample size over all loci.
[b] Proportion of polymorphic loci.
[c] Mean number of alleles per locus.
[d] Mean number of alleles per polymorphic locus.
[e] Expected heterozygosity.
[f] Observed heterozygosity.
[g] Fixation index.

supports their classification as hybrids. Almost every 'Green Malayan Dwarf' and 'Yellow Malayan Dwarf' individual and three 'Panama Talls', did not seem to have unique membership; instead they appeared as mixtures or hybrids of several cultivars. In the STRUCTURE analysis with the microsatellite data obtained by Meerow et al. (2003), eight populations ($K = 8$) were identified in all 110 genotypes (Fig. 1B). Cultivars identified as unique populations were 'Red Malayan Dwarf', 'Green Niño' and 'Red Spicata'. The 'Fiji Dwarf' cultivar was divided in two populations whereas the 'Green' and 'Yellow Malayan Dwarf' cultivars were identified as a single population. The 'Atlantic Tall' and 'Panama Tall' cultivars were highly heterogeneous and appeared largely as mixtures of hybrids. However, distinct allele frequencies were identified for the 'Panama Tall' cultivar (purple color almost unique to this cultivar, Fig. 1B). As with the WRKY results, 'Maypan' and the off-types displayed a hybrid composition.

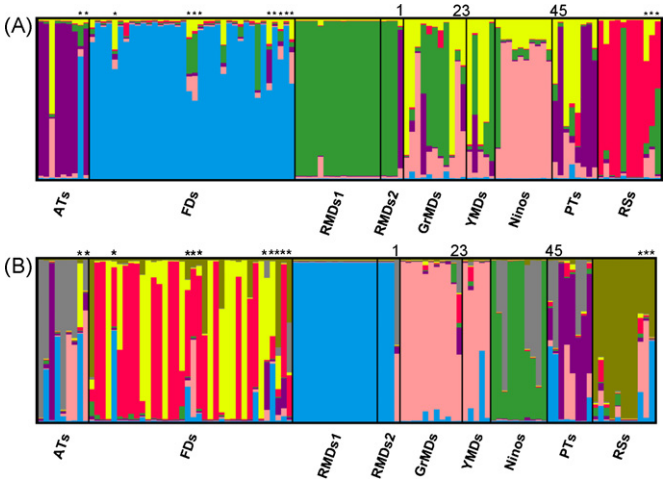

Fig. 1. (A) and (B) results from STRUCTURE analysis of WRKY and SSR data across 110 coconut genotypes representing eight *Cocos nucifera* cultivars and some hybrids. A. WRKY data. B. SSR data. Individuals marked with * are putative hybrids and outliers removed from further analyses. 1 = off-type 'Red Malayan Dwarf', 2, 3 = off-type 'Green Malayan Dwarfs', 4 = 'Maypan', 5 = Undetermined tall. ATs = 'Atlantic Tall', FDs = 'Fiji Dwarf', RMDs1 = 'Red Malayan Dwarf'-1, RMDs2 = 'Red Malayan Dwarf'-2, GrMDs = 'Green Malayan Dwarf', YMDs = 'Yellow Malayan Dwarf', Ninos = 'Green Niño', PTs = 'Panama Tall', RSs = 'Red Spicata'.

We identified a total of 37 alleles with all 13 WRKY-derived markers, ranging between two and four alleles per marker, and an average of 2.84 ± 0.69. For every marker the expected heterozygosity was higher than the observed, which resulted in fixation indices higher than 0.5 for all but four of the markers. Gene diversity (H_{sk}) ranged from 0.075 (CnWRKY-02, for which one of its two alleles was present only in few of the tall genotypes tested) to 0.542 (CnWRKY-05, with three alleles with variable frequencies across cultivars); the mean was 0.345 ± 0.168 (Table 2). The cultivar with the highest number of unique alleles was 'Fiji Dwarf', with six out of nine unique alleles identified. A unique allele was also identified for the cultivars 'Panama Tall', 'Red Spicata' and 'Yellow Malayan Dwarf' (Table 3). In addition, the overall analyses with the WRKY and the SSR data showed the same trend with each marker system: lower observed heterozygosity than the expected and equivalent fixation indexes of 0.46 (Table 4).

The overall cluster analysis with the neighbor-joining (NJ) method and Rogers' distance estimates for the same 93 genotypes used in Meerow et al. (2003), shows five distinct clusters: one for 'Atlantic Tall' and 'Panama Tall' genotypes, a cluster of 'Fiji Dwarf' genotypes, one of the 'Red Spicata' cultivar, two clusters for the 'Green Niño' samples and a heterogeneous cluster including the 'Red Malayan Dwarf' genotypes plus a few 'Green' and 'Yellow Malayan Dwarfs' (Fig. 2). In addition, most of the individuals that had a hybrid composition according to the STRUCTURE results with the WRKY data (e.g. several 'Panama Tall' and 'Green' and 'Yellow Malayan Dwarf' samples) were dispersed throughout the NJ dendrogram.

The two UPGMA phenograms of cultivar relationships obtained with either Nei's (1978) unbiased distance or modified Rogers' distance (Wright, 1978) had very similar tree topologies and bootstrap support values (Fig. 3). The nodes with the highest bootstrap support (>70%) linked the two sample sets for the 'Red Malayan Dwarf' cultivar (which had a genetic identity, $I = 1$, Table 5), the 'Green Malayan Dwarf' and 'Yellow Malayan Dwarf' cultivars (with $I = 1$ as well, Table 5), and the two tall cultivars with the cluster inclusive of the 'Green Niño' and all the 'Malayan Dwarf' cultivars. The genetic identities between each tall cultivar and the 'Malayan Dwarf' and 'Green Niño' cultivars were higher than 0.85 (Table 5). The node connecting the 'Atlantic Tall' and the 'Panama Tall' cultivars ($I = 0.957$, Table 5), and that for the 'Fiji Dwarf' and 'Red Spicata' cultivars ($I = 0.751$, Table 5) had bootstrap support between 40–50%.

The nodes with the lowest support values were those linking all the 'Malayan Dwarf' cultivars as well as the node linking the 'Malayan Dwarf' cultivars with the 'Green Niño' cultivar (bootstrap support between 30-47%). The 'Red Spicata' and 'Fiji Dwarf' cultivars formed a distinct cluster in both phenograms (Fig. 3). Accordingly, these two cultivars had the lowest average genetic identities with all other cultivars (0.699 and 0.726, respectively). The Mantel tests performed between the genetic distance matrices obtained with the WRKY data (used to generate these phenograms) and the microsatellite data (Meerow et al., 2003) gave a correlation of 0.586

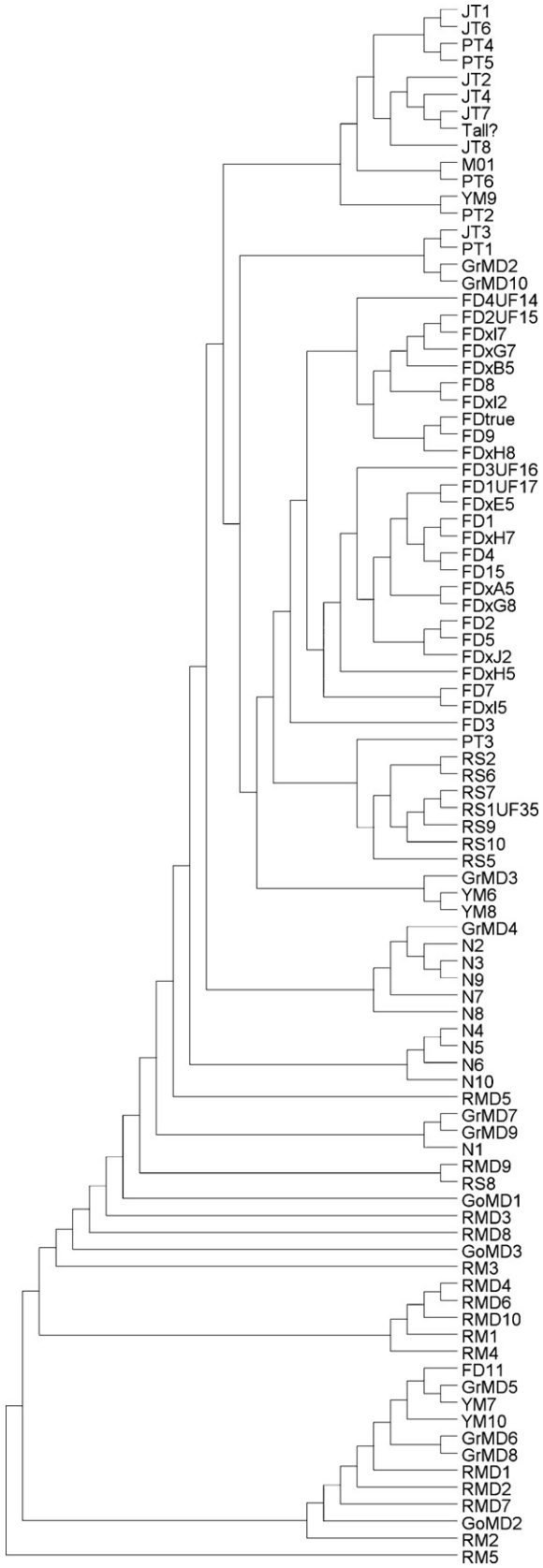

Fig. 2. Neighbor joining tree using modified Rogers distance (Wright, 1978) based on 13 WRKY-derived markers for 93 coconut genotypes.

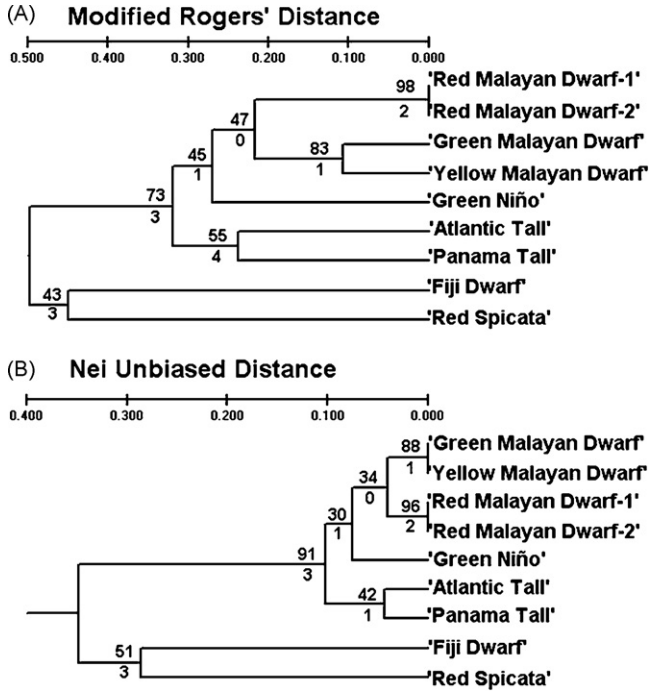

Fig. 3. UPGMA dendrograms for eight coconut cultivars using genetic distance measures from 13 WRKY loci. The number above the branch indicates the bootstrap percentage and the one below indicates the number of loci supporting that node. (A) Modified Rogers' distance (Wright, 1978). (B) Nei's (1978) unbiased distance.

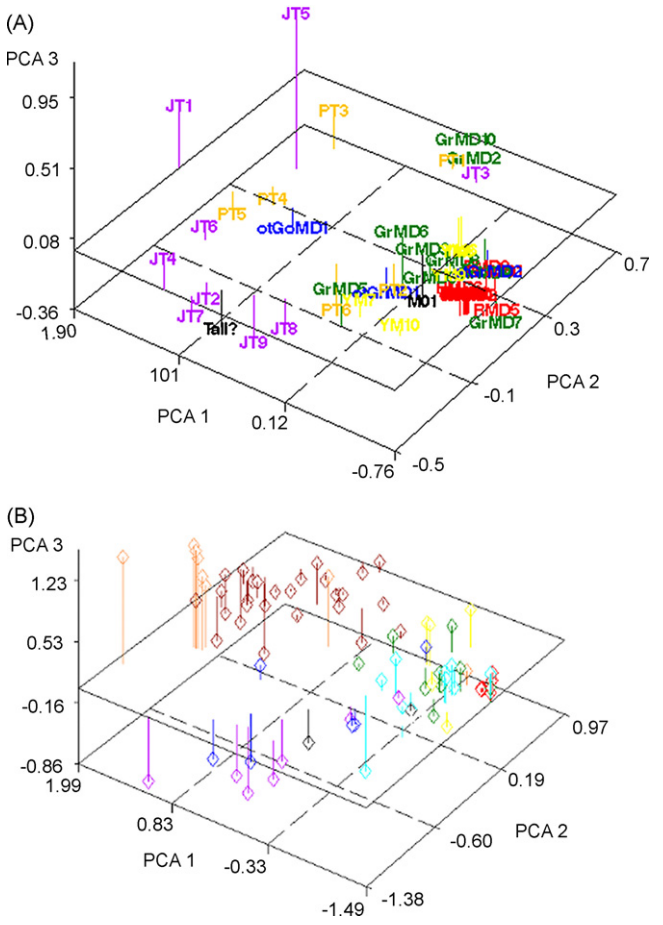

Fig. 4. (A) and (B) PCA plots of coconut genotypes with Modified Rogers' distances (Wright, 1978) using 13 WRKY-derived markers. A. 'Tall' and 'Malayan Dwarf' individuals. JT = 'Atlantic Tall', PT = 'Panama Tall', Tall? = Undetermined tall, M01 = 'Maypan', GrMD = 'Green Malayan Dwarf', RMD = 'Red Malayan Dwarf', YMD = 'Yellow Malayan Dwarf', otGoMD = off-type 'Red Malayan Dwarf'. PCA1 explains 75% of the variation, PCA2 explains 9.7% of the variation and PCA3 explains 6.8% of the variation; total is 91.5%. B. Ninety-three coconut genotypes (putative hybrids and outliers removed) 'Atlantic Tall' = purple, 'Panama Tall' = blue, 'Green Niño' = cyan, 'Green Malayan Dwarf' = green, 'Red Malayan Dwarf' = red, 'Yellow Malayan Dwarf' = yellow, 'Fiji Dwarf' = brown, 'Red Spicata' = salmon, undetermined tall = black. PCA1 explains 68.9% of the variation, PCA2 explains 11.2% of the variation and PCA3 explains 7.5% of the variation; total is 87.7%.

(Prob. $\leq$ 0.01) for the Nei's (1978) unbiased distances, and 0.668 (Prob. $\leq$ 0.001) for the modified Rogers' distances (Wright, 1978).

Two PCA analyses were performed. The first included the tall and 'Malayan Dwarf' cultivars, the 'Maypan' hybrid and several off-types, as in Meerow et al. (2003). The second included all eight cultivars (91 genotypes without off-types), the undetermined tall and the 'Maypan' hybrid. In the first analysis, two main groups can be identified in the PCA plot, one including primarily 'Atlantic Tall' genotypes and the group for the 'Malayan Dwarf' cultivars. Each one of these groups also contains some 'Panama Talls' and off-types (Fig. 4A). Some individuals were placed in between these two groups, e.g., PT6, and may therefore be of hybrid origin. The cumulative variation explained by the three axes of the PCA plot was 91.5%. In the

Table 5
Pairwise Nei's (1978) unbiased genetic identity (above diagonal) and modified Rogers' genetic distance (Wright's 1978, below diagonal) between the coconut cultivars evaluated. Refer to Fig. 4 for cultivar identification

	AT	FD	RMD1	RMD2	GrMD	YMD	N	PT	RS
AT		0.673	0.858	0.858	0.918	0.921	0.883	0.957	0.607
FD	0.486		0.687	0.687	0.775	0.786	0.703	0.750	0.751
RMD1	0.383	0.527		1.000	0.966	0.955	0.904	0.868	0.662
RMD2	0.383	0.527	0.000		0.966	0.955	0.904	0.868	0.662
GrMD	0.289	0.425	0.206	0.206		1.000	0.962	0.960	0.741
YMD	0.294	0.424	0.227	0.227	0.107		0.940	0.960	0.758
N	0.341	0.497	0.309	0.309	0.207	0.252		0.940	0.692
PT	0.238	0.431	0.375	0.375	0.236	0.243	0.272		0.726
RS	0.563	0.449	0.565	0.565	0.475	0.467	0.526	0.479	

PCA that included all eight cultivars, four groupings can be clearly distinguished: one each for the 'Red Spicata', the 'Fiji Dwarf', the 'Red Malayan Dwarf' and the talls (primarily 'Atlantic Talls') (Fig. 4B). The 'Green' and 'Yellow Malayan Dwarf' samples, plus the 'Green Niño' were grouped together. The 'Panama Talls' were placed in between groups or as part of one of them. The cumulative variation explained by the three axes of the PCA plot was 87.7%.

4. Discussion

4.1. Informativeness of WRKY-derived markers as compared to microsatellite markers

Despite the lower number of alleles identified by the 13 WRKY-derived markers in contrast with the 15 microsatellite markers (37 versus 67 total alleles or 2.84 ± 0.69 versus 4.31 ± 1.8 alleles per locus), there was a good level of correspondence between the results of each marker type. Likewise, per locus average gene diversity was 0.345 ± 0.168 for the WRKY markers versus 0.574 ± 0.140 for the SSR markers. For both marker systems the expected heterozygosity was almost twice the observed one, leading to almost identical fixation indexes (0.465 versus 0.46, Table 4). This last result reflects the higher number of self-pollinated genotypes included in this study. Regarding unique alleles, both WRKY and microsatellite studies highlighted 'Fiji Dwarf' as the cultivar with the most unique alleles (six and seven unique alleles, respectively). In the microsatellite study (Meerow et al. (2003)), the 'Atlantic Tall' cultivar had two unique alleles and the 'Green Niño' had one. In the WRKY study, 'Panama Tall', 'Red Spicata' and 'Yellow Malayan Dwarf' cultivars each had a single unique allele (Table 3). The informativeness of the WRKY-derived markers, despite their lower allele number, seems to derive from their allelic distribution across cultivars. For example, with CnWRKY-01 (with three alleles identified) one allele is present only in the 'Red Spicatas' and 'Fiji Dwarfs'; a second allele is present only in the 'Fiji Dwarfs', while all the other cultivars are homozygous for the third allele (data not shown).

4.2. Identification of cultivars, hybrids and off-types with each marker type

General similarities and also some differences were identified in the genetic population structure of the 110 coconut genotypes as resolved with each marker type. In general, the clustering results of each with STRUCTURE (Fig. 1A and B) matched their corresponding NJ clustering tree results (Fig. 2 this paper and Fig. 1 in Meerow et al. (2003)) and PCA results (Fig. 4A and B this paper and Fig. 3 in Meerow et al. (2003)). The STRUCTURE analysis with the microsatellite data detected eight populations versus six populations detected with the WRKY data, which may in part be due to the limited polymorphism of SNP markers compared to microsatellites. Both marker types identified the 'Red Malayan Dwarf' and 'Red Spicata' cultivars as distinct populations with every clustering method. The same could be said of the 'Fiji Dwarf' and 'Green Niño' cultivars with the exception that STRUCTURE identified two groups within the 'Fiji Dwarf' cultivar (Fig. 1B) with the microsatellite data, while the PCA analysis with the WRKY data placed the 'Green Niño' with the 'Malayan Dwarf' cultivars (Fig. 4B). These slight differences indicate variations in the clustering methods as well as in the ability of each molecular marker type to discriminate cultivars from one another.

Both marker types also grouped together most of the 'Atlantic Tall' and 'Panama Tall' individuals across clustering methods, but some differences are evident between microsatellite and WRKY markers. The STRUCTURE analysis with the microsatellite data resolved the 'Atlantic' and 'Panama Talls' as mixtures of admixed/hybrid individuals between these two cultivars or with the 'Malayan Dwarfs' (Fig. 1B), a result congruent with their outcrossing behavior. In addition, in the NJ cluster tree (microsatellite data, Fig. 1 in Meerow et al. (2003)) all the tall individuals were grouped together except for those that according to STRUCTURE had some 'Red Malayan Dwarf' background (blue color, Fig. 1B). These individuals were placed close to the 'Malayan Dwarf' cluster. On the other hand, the STRUCTURE analysis with the WRKY data showed most of the tall genotypes as a single population (purple color, Fig. 1A), which also grouped together in the WRKY PCA analysis (Fig. 4 A and B) and NJ cluster tree (Fig. 2). The remaining three 'Panama Talls' and JT3 (an 'Atlantic Tall') were genotypically aligned with some 'Green' or 'Yellow Malayan Dwarfs'. Whether as a mixture of hybrids or as a more homogeneous group, the two tall cultivars were grouped together with both marker systems. This situation is being further investigated since these two cultivars are supposed to have completely different origins (Lebrun et al., 2005).

In regards to the 'Green' and 'Yellow Malayan Dwarf' cultivars, the microsatellite data identified them as a single population (Meerow et al., 2003). However, the WRKY data portrayed these two cultivars as highly heterogeneous (Fig. 1A), and several of their individuals grouped with or in between other cultivars, except in the PCA where the two cultivars clustered together (Fig. 4 A and B). The observed differences between SSR and WRKY results may be attributed to intrinsic differences in each marker system. The microsatellite markers were developed by random cloning (Perera et al., 1999), so we presume they may be neutral and located in non-coding regions. Conversely, some of the WRKY loci may not be neutral markers since many WRKY genes are involved in functions essential to the plant (Ulker and Somssich, 2004; Xie et al., 2005). Whether we have WRKY markers that have a selective advantage (such as resistance to some diseases) requires further research.

Lastly, the 17 individuals removed from the microsatellite analysis after being identified as putative hybrids (Meerow et al., 2003), were also identified in the STRUCTURE analyses of both the microsatellite and the WRKY data as possible hybrids (Fig. 1A and B). Furthermore, other individuals not removed in Meerow et al. (2003) also showed some level of admixture with both marker types.

4.3. Relationships between cultivars

Two clusters were consistent across the two WRKY dendrograms (Fig. 3) and the two microsatellite dendrograms (Fig. 2A and B in Meerow et al. (2003)). The first includes the three 'Malayan Dwarf' cultivars and the second unites the 'Atlantic' and 'Panama' tall cultivars. The 'Malayan Dwarf' cluster is congruent with common provenance of Java via Malaysia for these varieties (Harries, 1978). The grouping of the two tall cultivars across all four dendrograms is puzzling since they theoretically originated in different groups: the Pacific and Indo-Atlantic (Lebrun et al., 2005). These two tall cultivars had high genetic identity (0.957, Table 5), and the STRUCTURE results also indicate high similarity between them (Fig. 1A and B). Extensive hybridization in Florida between these two outcrossing cultivars was also suggested in Meerow et al. (2003). Whether the clustering of these two tall cultivars was due primarily to high levels of hybridization and not to a common origin needs further investigation.

The link between the cluster of the two tall cultivars and any of the other cultivars was not supported by the microsatellite genetic data (Fig. 2A and B in Meerow et al. (2003)). However, the two WRKY dendrograms indicate a well-supported linkage with the 'Green Niño'-'Malayan Dwarf' cluster (Fig. 3). The 'Green Niño', a dwarf cultivar (Harries, 1978), did not have a supported position in all four dendrograms (two for WRKY data and two for microsatellite data) but in the WRKY dendrograms it was consistently clustered with the three 'Malayan Dwarf' cultivars. Dwarf cultivars, according to the provisional classification of coconut cultivars (Lebrun et al., 2005), should belong to the Pacific group. Furthermore, the two STRUCTURE results (Fig. 1A and B) show some genetic similarities among individuals of these six cultivars, i.e. the two tall cultivars, the 'Green Niño' and the three 'Malayan Dwarf' cultivars. Lebrun et al. (1998), Teulat et al. (2000) and Perera et al. (2003), also placed the 'Panama Tall' and 'Malayan Dwarf' cultivars in the same general group since they are presumed to share a common ancestry within the Pacific group (Lebrun et al., 2005). Our results may indicate that the high level of hybridization between the two tall cultivars resulted in the 'Atlantic Talls' of our study resolving with the 'Panama Tall'-'Malayan Dwarf'-'Green Niño' (i.e. Pacific group) cluster, or possibly that the 'Atlantic Talls' were mislabeled, all of which is being further investigated.

The 'Fiji Dwarf' and 'Red Spicata' cultivars grouped together in one of the microsatellite dendrograms (Fig. 2A and B in Meerow et al. (2003)) and in the two WRKY-derived dendrograms (Fig. 3), though the bootstrap support with the WRKY data was weak (43 and 51%). This result is perplexing since the 'Fiji Dwarf' cultivar has been grouped with the Pacific group in other studies (Lebrun et al., 1998; Teulat et al., 2000; Perera et al., 2003). If our 'Atlantic Tall' samples are hybrids or mislabeled, then 'Red Spicata' is the only cultivar we included from the Indo-Atlantic group. The microsatellite dendrograms resolved an ambiguous relationship for these two cultivars with the others. Furthermore, these two cultivars had the lowest genetic identities with the other cultivars with both the microsatellite (Table 8 in Meerow et al. (2003)) and the WRKY data (Table 5). It is unclear why the 'Fiji Dwarf' and 'Red Spicata' cultivars were clustered together and had similar genetic relationships with the other cultivars. Again, it might be a result of the type of variability/information revealed by the WRKY markers. Nevertheless, a broader representation of cultivars from the two main groups might better resolve the relationships among them.

The comparison of results obtained with different marker types has provided similar results (Lebrun et al., 1998), or slight differences between marker systems. For example, Teulat et al. (2000) found AFLP markers were better than SSRs at assessing genetic relationships between cultivars. The Mantel tests for genetic distance correlation between cultivars for WRKY and SSR marker systems was 0.586 ($p \leq 0.01$) for Nei's (1978) unbiased distances, and 0.668 ($p \leq 0.001$) for modified Rogers' distances (Wright, 1978), indicating at best a moderate correspondence between the information generated by SSRs and WRKY-derived markers.

In conclusion, WRKY-derived markers proved useful in the assessment of the population structure of several coconut cultivars. WRKY-derived markers can complement information already generated with other markers, as well as reveal patterns of genetic diversity undetectable with microsatellite markers. WRKY markers may have the advantage of being linked to phenotypic traits of interest, which could make them very useful in QTL analysis.

Acknowledgement

We would like to thank Robert Sautter for help with the PCA graphics.

References

Ashburner, G.R., Thompson, W.K., Halloran, G.M., 1997. RAPD analysis of South Pacific coconut palm populations. Crop Sci. 37, 992–997.

Borrone, J.W., Kuhn, D.N., Schnell, R.J., 2004. Isolation, characterization, and development of WRKY genes as useful genetic markers in *Theobroma cacao*. Theor. Appl. Genet. 109, 495–507.

Dong, J., Chen, C., Chen, Z., 2003. Expression profiles of the *Arabidopsis* WRKY gene superfamily during plant defense response. Plant Mol. Biol. 51, 21–37.

Eulgem, T., Rushton, P.J., Schmelzer, E., Hahlbrock, K., Somssich, I.E., 1999. Early nuclear events in plant defense signaling: rapid gene activation by WRKY transcription factors. EMBO J. 18, 4689–4699.

Eulgem, T., Rushton, P.J., Robatzek, S., Somssich, I.E., 2000. The WRKY superfamily of plant transcription factors. Trends Plant Sci. 5, 199–206.

Falush, D., Stephens, M., Pritchard, J.K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164, 1567–1587.

Foale, M., 2005. An introduction to the coconut palm. In: Batugal, et al. (Eds.), Coconut Genetic Resources.. International Plant Genetic Resources Institute, Regional Office for Asia, the Pacific and Oceania (IPGRI-APO), Serdang, Selangor DE, Malaysia.

Glaubitz, J.C., 2004. CONVERT: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. Mol. Ecol. Notes 4, 309–310.

Goudet, J., 2002. FSTAT (vers. 2.9.3.2), a program to estimate and test gene diversities and fixation indices. Available from: http://www2.unil.ch/popgen/softwares/fstat.htm.

Harries, H.C., 1978. The evolution, dissemination and classification of *Cocos nucifera* L. Bot. Rev. 44, 265–320.

Harries, H., Baudouin, L., Cardeña, R., 2004. Floating, boating and introgression: molecular techniques and the ancestry of coconut palm populations on Pacific islands. Ethnobot. Res. Appl. 2, 37–53. (online)In: http://www.ethnobotanyjournal.org/vol2/i1547-3465-02-037.pdf.

Johnson, C.S., Kolevski, B., Smyth, D.R., 2002. TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. Plant Cell 14, 1359–1375.

Kuhn, D.N., Borrone, J., Meerow, A.W., Motamayor, J.C., Brown, J.S., Schnell, R.J., 2005. Single-strand conformation polymorphism analysis of candidate genes for reliable identification of alleles by capillary array electrophoresis. Electrophoresis 26, 112–125.

Langella, O., 2002. Populations 1.2.28: Population genetic software (individuals or populations distances, phylogenetic trees). http://www.cnrs-gif.fr/pge.

Lebrun, P., N'cho, Y.P., Seguin, M., Grivet, L., Baudouin, L., 1998. Genetic diversity in coconut (*Cocos nucifera* L.) revealed by restriction fragment length polymorphism (RFLP) markers. Euphytica 101, 103–108.

Lebrun, P., Berger, A., Hodgkin, T., Baudouin, L., 2005. Biochemical and molecular methods for characterizing coconut diversity. In: Batugal, et al. (Eds.), Coconut Genetic Resources. International Plant Genetic Resources Institute, Regional Office for Asia, the Pacific and Oceania (IPGRI-APO), Serdang, Selangor DE, Malaysia.

Lewis, P.O., Zaykin, D., 2001. GENETIC DATA ANALYSIS: computer program for the analysis of allelic data. Version 1.1. Free program distributed by the authors over the internet from http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php.

Liyanage, D.V., 1949. Preliminary studies on the floral biology of the coconut palm. Trop Agric 105, 171–175.

Luo, M., Dennis, E.S., Berger, F., Peacock, W.J., Chaudhury, A., 2005. MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in *Arabidopsis*. PNAS 102, 17531–17536.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27, 209–220.

Mauro-Herrera, M., Meerow, A.W., Borrone, J.W., Kuhn, D.N., Schnell, R.J., 2006. Ten informative markers developed from WRKY sequences in coconut (*Cocos nucifera*). Mol. Ecol. Notes 6, 904–906.

Meerow, A.W., Wisser, R.J., Brown, J.S., Kuhn, D.N., Schnell, R.J., Broschat, T.K., 2003. Analysis of genetic diversity and population structure within Florida coconut (*Cocos nucifera* L.) germplasm using microsatellite DNA, with special emphasis on the Fiji Dwarf cultivar. Theor. Appl. Genet. 106, 715–726.

Miller, M.P., 1997. Tools for population genetic analyses (TFPGA) 1.3: A windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by author. http://www.marksgeneticsoftware.net/.

Nei, M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89, 583–590.

Nei, M., 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Perera, L., Russell, J.R., Provan, J., McNicol, J.W., Powell, W., 1998. Evaluating genetic relationships between indigenous coconut (*Cocos nucifera* L.) accessions from Sri Lanka by means of AFLP profiling. Theor. Appl. Genet. 96, 545–550.

Perera, L., Russell, J.R., Provan, J., Powell, W., 1999. Identification and characterization of microsatellite loci in coconut (*Cocos nucifera* L.) and the analysis of coconut populations in Sri Lanka. Mol. Ecol. 8, 344–346.

Perera, L., Russell, J.R., Provan, J., Powell, W., 2003. Studying genetic relationships among coconut varieties/populations using microsatellite markers. Euphytica 132, 121–128.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Rao, V.R., Hodgkin, T., Bourdeix, R., 2005. Locating coconut genetic diversity. In: Batugal, et al. (Eds.), Coconut Genetic Resources.. International Plant Genetic Resources Institute, Regional Office for Asia, the Pacific and Oceania (IPGRI-APO), Serdang, Selangor DE, Malaysia.

Rizhsky, L., Liang, H., Mittler, R., 2002. The combined effect of drought stress and heat shock on gene expression in tobacco. Plant Physiol. 130, 1143–1151.

Robatzek, S., Somssich, I.E., 2001. A new member of the Arabidopsis WRKY transcription family, AtWRKY6, is associated with both senescence- and defense-related processes. Plant J. 28, 123–133.

Rohde, W., Kullaya, A., Rodriguez, J., Ritter, E., 1995. Genome analysis of *Cocos nucifera* L. by PCR amplification of spacer sequences separating a subset of *Copia*-like *Eco*R1 repetitive elements. J. Genet. Breed. 49, 179–186.

Rosenberg, N.A., 2004. DISTRUCT: a program for the graphical display of population structure. Mol. Ecol. Notes 4, 137–138.

Ryu, H.S., Han, M., Lee, S.K., Cho, J.I., Ryoo, N., Heu, S., Lee, Y.H., Bhoo, S.H., Wang, G.L., Hahn, T.R., Jeon, J.S., 2006. A comprehensive expression analysis of the WRKY gene superfamily in rice plants during defense response. Plant Cell Rep. 25, 836–847.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Sneath, P.H.A., Sokal, R.R., 1973. Numerical Taxonomy: The Principles and Practice of Numerical Classification. WH Freeman, San Francisco.

Teulat, B., Aldam, C., Trehin, R., Lebrun, P., Barker, J.H.A., Arnold, G.M., Karp, A., Baudouin, L., Rognon, F., 2000. An analysis of genetic diversity in coconut (*Cocos nucifera*) populations from across the geographic range using sequence-tagged microsatellites (SSRs) and AFLPs. Theor. Appl. Genet. 100, 764–771.

Ulker, B., Somssich, I.E., 2004. WRKY transcription factors: from DNA binding towards biological function. Curr. Opin. Plant Biol. 7, 491–498.

Wright, S., 1978. Evolution and the Genetics of Populations, Variability within and Among Natural Populations, vol.4. University of Chicago Press, Chicago.

Xie, Z., Zhang, Z.L., Zou, X., Huang, J., Ruas, P., Thompson, D., Shen, Q.J., 2005. Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. Plant Physiol. 137, 176–189.

Yu, D., Chen, C., Chen, Z., 2001. Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression. Plant Cell 13, 1527–1539.

Zhang, Y., Wang, L., 2005. The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. BMC Evol. Biol. 5, 1–12.